Obtaining population-wide inequality estimates for Switzerland by reweighting high-quality subpopulation tax data

Rudolf Farys, Oliver Hümbelin, and Ben Jann

University of Bern and Bern University of Applied Sciences

Erster Kongress der Akademie für Soziologie
Munich, April 4–6, 2018

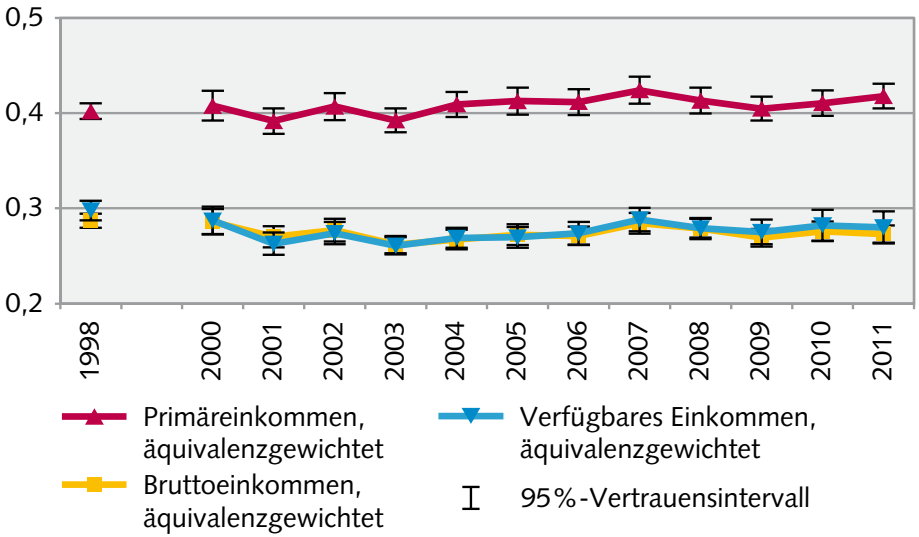
Contents

Inequality in Switzerland

- Income inequality increased in many OECD countries (OECD 2008, 2011, 2015; Salverda et al. 2014) .
- Switzerland: Results are ambiguous, depending on data source
  - Survey-based estimates indicate stable, or even decreasing inequality (e.g. Household Budget Survey).
  - Tax-data-based estimates indicate increasing inequality, in particular at the upper end of the distribution.

Inequality in Switzerland

Entwicklung der Gini-Koeffizienten 1998 bis 2011, Gesamtbevölkerung[1] **G 6**

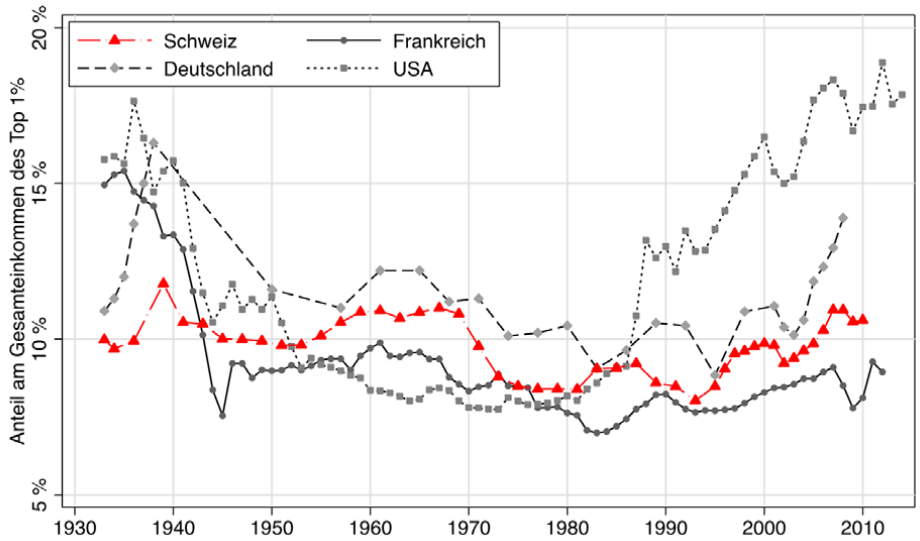


[1] Berechnungen einschliesslich der negativen Einkommen

Quelle: Haushaltsbudgeterhebung © *BFS*

(Bundesamt für Statistik 2013)

Grafik 1. Einkommensanteil des Top 1% im internationalen Vergleich



(Martinez 2017)

Survey data vs. tax data

- Survey data:
  - Pros:
    - ★ measurement of income based on theory-guided income definitions
  - Cons:
    - ★ strong middle-class bias, underrepresentation of the top and the bottom
    - ★ small sample sizes
- Tax data:
  - Pros:
    - ★ full census
    - ★ income and assets in great detail
  - Cons:
    - ★ measurement of income for administrative purposes; some components lacking (e.g. social assistance)
    - ★ tax subjects, not households
    - ★ little additional information
    - ★ difficult to obtain from all cantons of Switzerland

Goal of this study

- We have access to high quality individual-level tax data from the canton of Bern. One of the advantages of this particular dataset is that households are identified.
- Can we use this dataset to estimate the level of inequality (in equivalized disposable income) in whole Switzerland?

Combination of different data sources

- Detailed tax data from canton of Bern (2012).
- Public tax statistics from the Federal Tax Administration (FTA).
- Various statistical indicators at the municipality level from the Federal Statistical Office (FSO).

Bern tax data

- Good:
  - Enough details to construct desirable income measures (disposable income).
  - Individual level-data; (nearly) full census.
  - Linked with Federal Register of Buildings and Dwellings to identify household structures.
- Bad:
  - Only one canton (about 12% of the Swiss population)

Tax statistics by the FTA

- Good:
  - Covers the complete population of Switzerland
  - Indicators (such as averages, medians, and Gini coefficients) at the municipality level
- Bad:
  - Inappropriate income measurement (taxable income)
  - Statistics are for "tax subjects", not households or individuals
    - ★ e.g. married couple = 1 tax subject; unmarried couple = 2 tax subjects
  - Aggregate data

Idea

- Use the FTA indicators (as well as other indicators from the FSO) to derive weights that can be applied to the tax data from Bern.
- The weights are constructed at the municipality level. The goal is to reweight the Bernese municipalities such that they look like Switzerland.
- These weights can then be used when analyzing the individual-level tax data from Bern.
- The procedure should work if there is enough heterogeneity among Bernese municipalities and if strong predictors for inequality at the municipality level are available.

Method

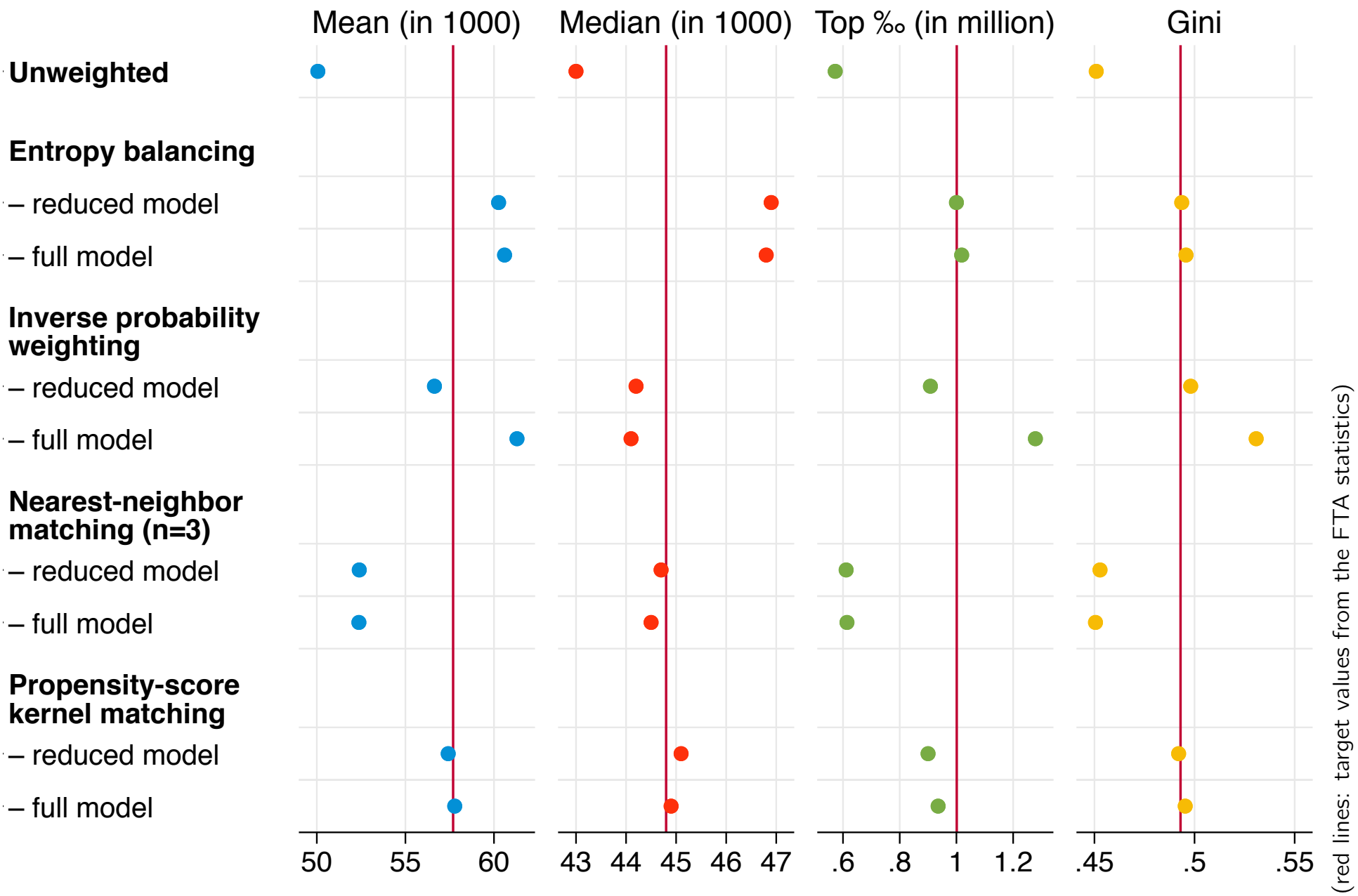
- We use various methods to compute the weights
  - Entropy balancing (Hainmüller 2012)
  - Inverse probability weighting
  - Nearest-neighbor matching
  - Propensity-score kernel matching (Jann 2017)

- Included variables (at the municipality level)
  - FTA indicators: average (pseudo-equivalized) taxable income, Gini coefficient of (pseudo-equivalized) taxable income
  + population size, distribution of household sizes, age distribution (reduced model)
  + economic structure of work force, proportion of welfare recipients (full model)

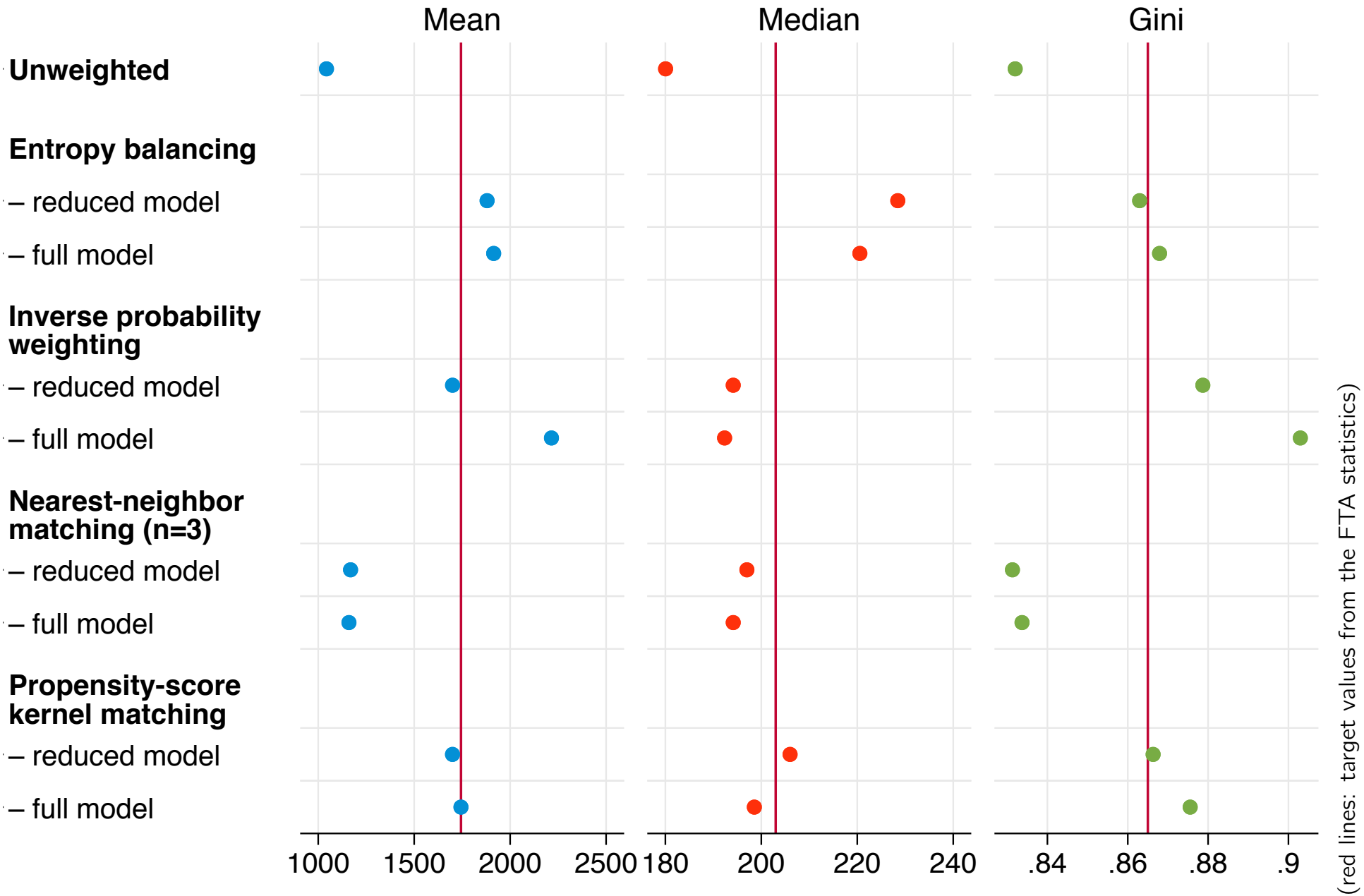
Results

- In the following we first evaluate how the different methods perform.
- We use the distribution of (non-equivalized) taxable income as well as the distribution of federal taxes paid by Swiss tax subjects as benchmark.
- That is, we evaluate whether these FTA statistics can be successfully reproduced by the reweighted Bernese tax data.
- We then provide estimates of the inequality of equivalized disposable income in Switzerland based on the reweighted data.

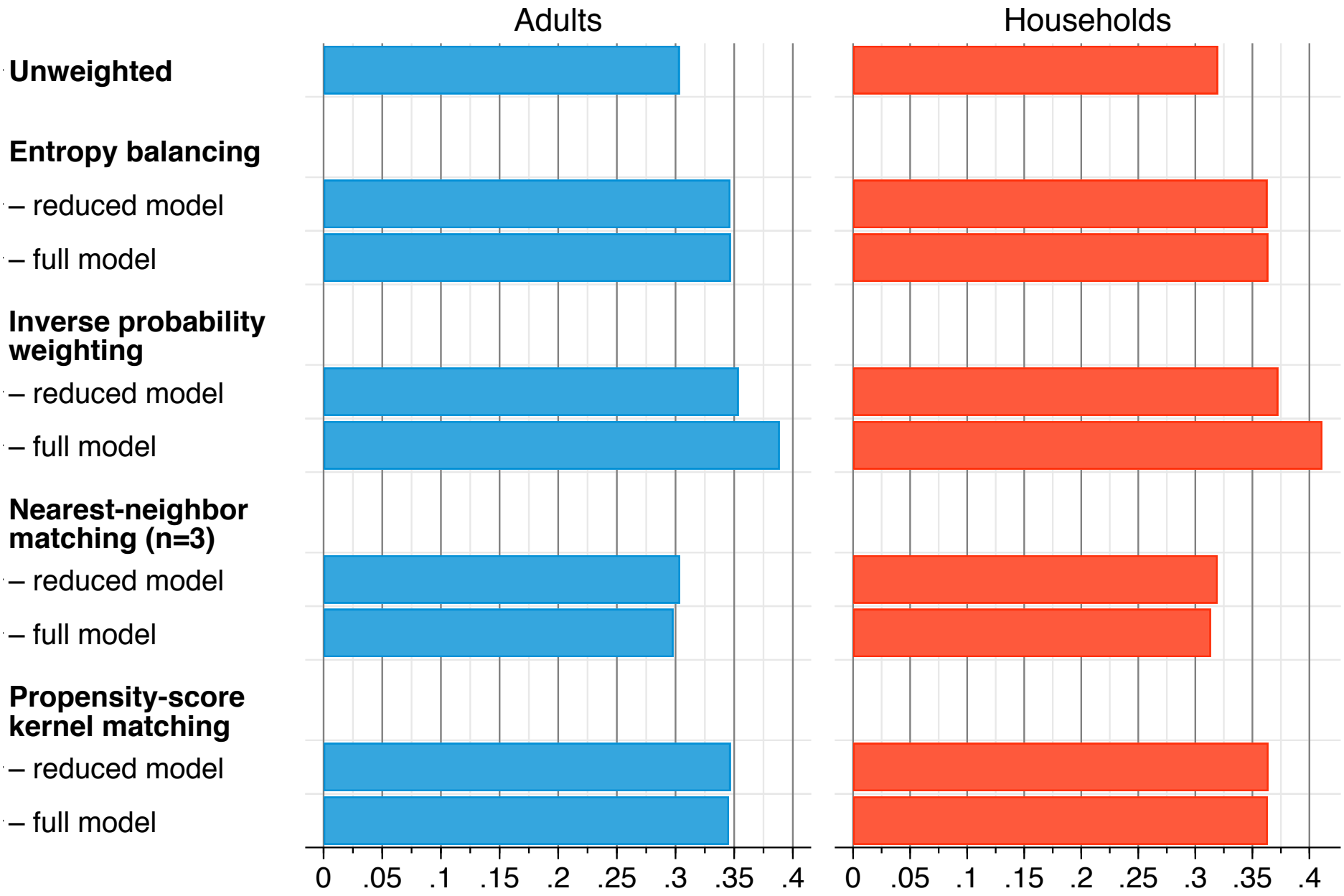
Taxable income



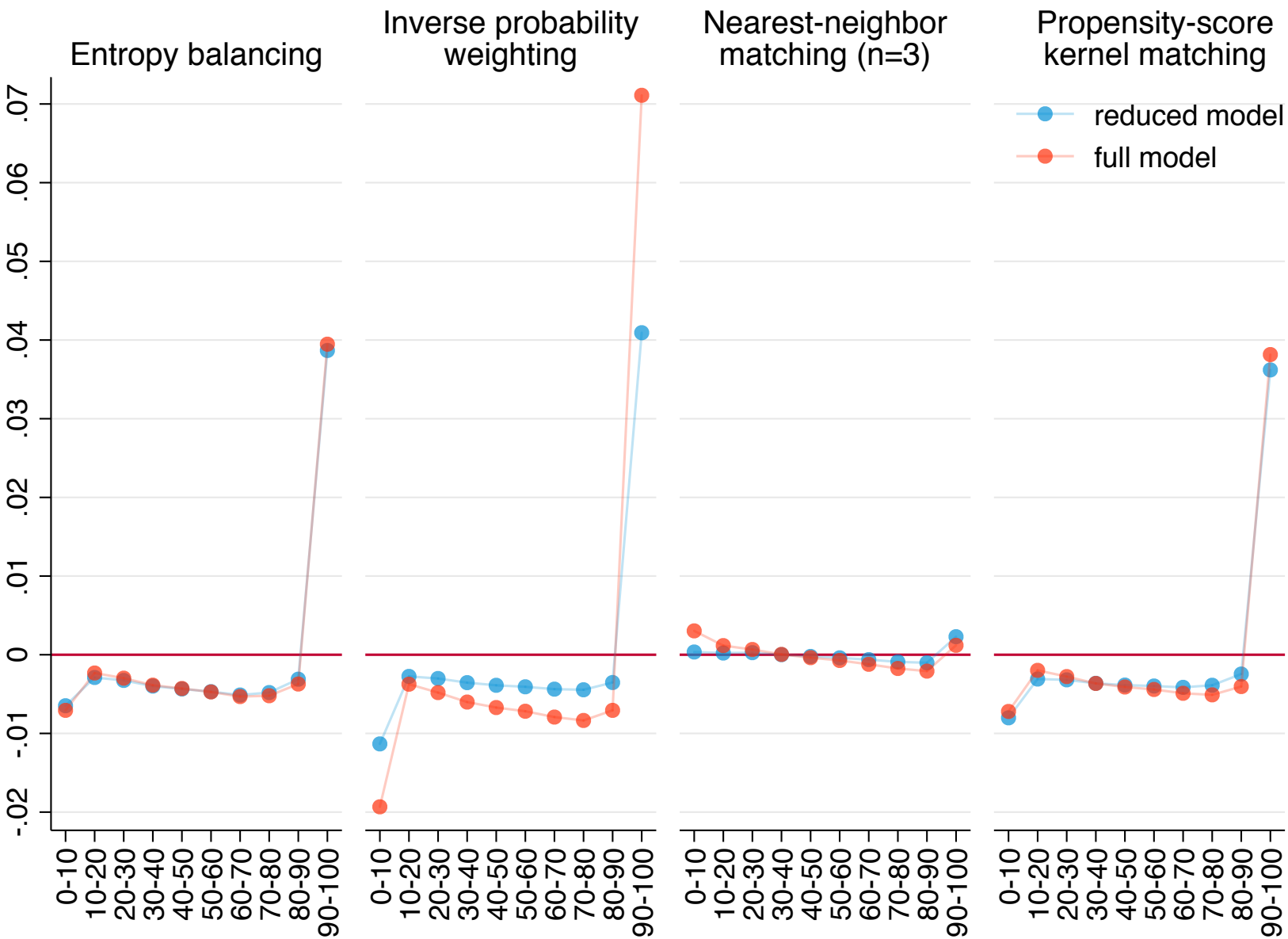
Tax bill (federal tax)



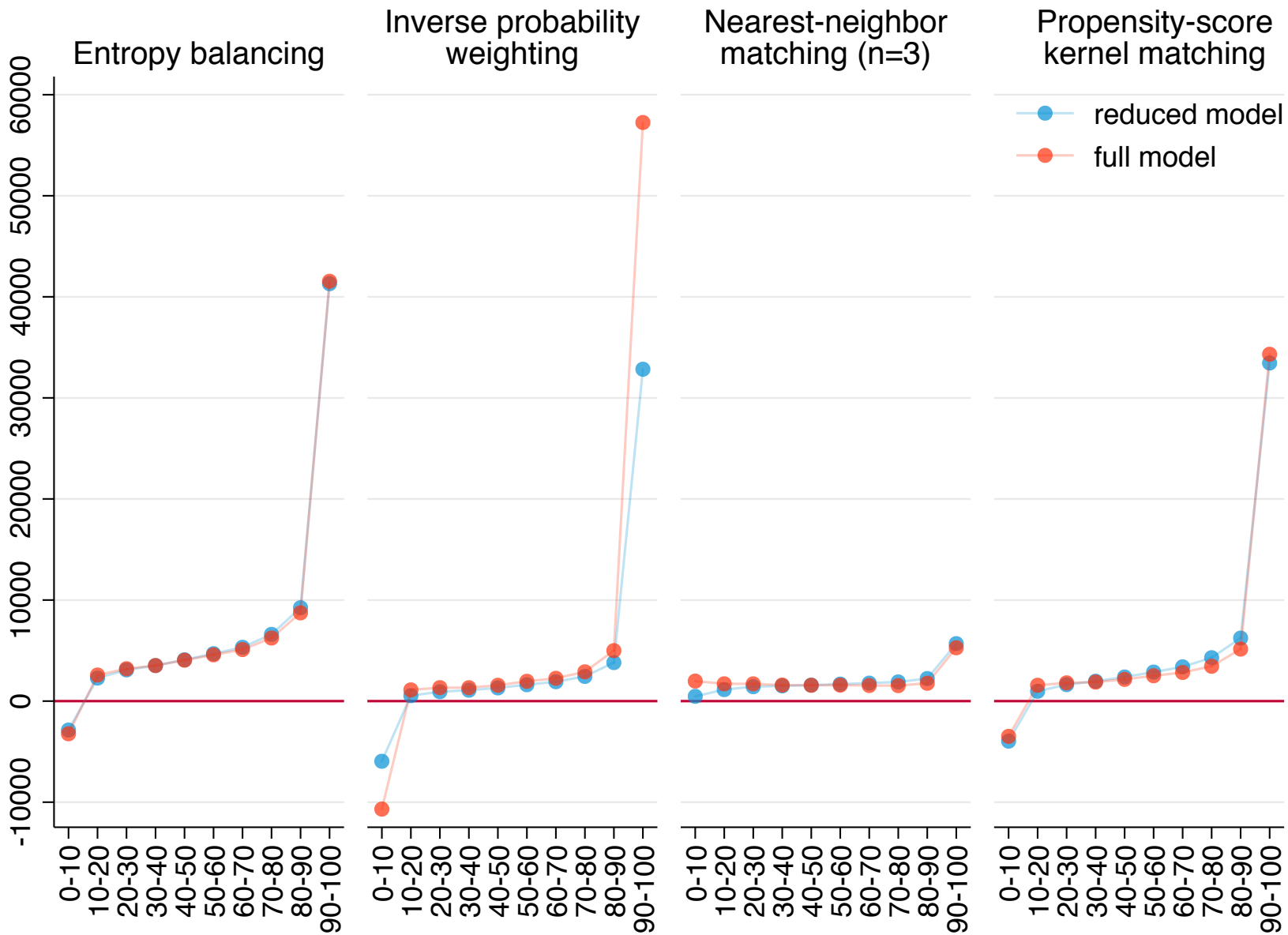
Gini (equivalized disposable income)



Difference in percentile shares between raw and reweighted data (equivalized disposable income):



Difference in average equivalized disposable income by percentile group between raw and reweighted data:



Conclusions

- The approach seems promising as the benchmark statistics could successfully be reproduced.
- However, the details of the procedure to compute the weights matter: Entropy balancing and propensity-score kernel matching were successful, nearest-neighbor matching and inverse probability weighting were not.
- The resulting Swiss Gini coefficient of equivalized disposable income is substantially higher than suggested by comparable survey based analyses (about .35 versus less than .30).

Outlook

- The tax data from Bern are not perfect. For example, information on social assistance is missing and the analytic potential is limited due to lack of interesting covariates (say, education).
- We just received a grant last week for a new project.
- In this project, cantonal tax records will be linked with . . .
  - . . . the population registry
  - . . . social security data
  - . . . the Swiss structural surveys (yearly surveys of about 250'000 residents on topic such as household and family, employment, education, etc.)
  - . . . several further administrative datasets and surveys
- The new database will allow accurate analyses of the economic situation (including assets) of households and individuals, and it will come with a rich set of covariates.

References

- Bundesamt für Statistik (2013). BFS Aktuell – Einkommen der privaten Haushalte: Einkommensungleichheit wird durch Umverteilung deutlich verringert. Neuchâtel: BFS.
- Hainmueller, J. (2012). Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies. Political Analysis (Winter 2012) 20 (1), 25-46.
- Jann, B. (2017). KMATCH: Stata module for multivariate-distance and propensity-score matching. Statistical Software Components S458346, Boston College Department of Economics.
- Martinez, I. (2017), Die Topeinkommen in der Schweiz seit 1980: Verteilung und Mobilität. Social Change in Switzerland No. 11. Retrieved from `http://socialchangeswitzerland.ch`
- OECD. (2008). Growing Unequal? Income Distribution and Poverty in OECD Countries. Paris: OECD Publishing.
- OECD (2011). Divided We Stand. Why Inequality Keeps Rising. Paris: OECD Publishing.
- OECD. (2015). In It Together: Why Less Inequality Benefits All. Paris: OECD Publishing.
- Salverda, W., Nolan, B., Checchi, D., Marx, I., McKnight, A. & Toth, I. G. (2014). Changing Inequalities in Rich Countries: Analytical and Comparative Perspectives. Oxford: Oxford University Press.